Deconstruction of the negative social heritage? – a search for variables confounding the simple relation between socioeconomic status and student achievement

Peter Allerup, Rune Müller Kristensen, André Torre

Aarhus University, Danish School of Education – Department of Educational Sociology

Key words: Negative social Heritage, log linear modelling, graphical modelling, conditional independence, confounding variables.

Abstract

Negative social heritage was earlier often addressed through studies in which parents' educational level were linked to the educational level of their children. Today this phenomenon is addressed through a statistical relation between children's achievement and parents' socioeconomic status, the so-called ESCS -measure (*E*conomic, *S*ocial and *C*ultural *S*tatus). The observed positive correlation between achievement and ESCS leads to a straight forward conclusion: Higher levels of socioeconomic status is associated with higher levels of attainment – and visa versa with low values. The reason for calling this relation 'negative social heritage', however is due to an early sociological study in which only low status ('negative') families participated. The relation between achievement and ESCS has been known for years, but it was PISA 2003 by focusing on the numerical level of correlation – 'weak' or 'strong' negative social heritage – who brought considerable attention to the field in Denmark by comparing the values across countries. It was clear that Denmark ranked among the countries having the highest values of negative social heritage. The numerous ways of calculating an adequate estimate of ESCS will be discussed in the paper. The common reference for the ESCS-measure is variables concerning parents' level of income, education and occupational status. Consequently, an immediate interpretation of the level of correlation between student achievement and ESCS is that part of the student's level of attainment is due to the family background. The aim of the paper is to disclose what is in fact controlling this correlation, as it is clear that whatever definition you may choose for the calculation of ESCS, no one believe that the attainment level of the student can be derived operationally from background variables for the parents.

TIMSS 2015 data has been the platform for exercising these analyses by involving all relevant student- and parent variables, from 'motivation' and 'confidence' to 'help to homework' etc. The statistical analysis technique is log linear modelling, more specifically graphical modelling, analyzing higher order contingency tables (grouped variables) for main effects and interaction effects in the light of conditional independence graphs. By this, it is hoped to reveal variables that may act as confounders for the simple correlation between student attainment and ESCS. The results, however, point to a conclusion, that none of the variables available can eliminate or diminish the basic correlation and one is consequently left with an open question: Are the variables determined by the TIMSS Framework comprehensible for such analyses?

Introduction

In the PISA 2003 Technical Report (OECD, 2004) the negative social heritage was presented by the graph in figure 1:

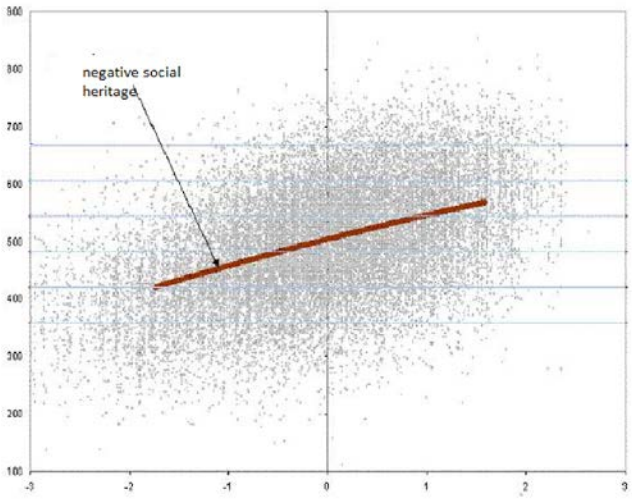

Figure 1: The relation between socioeconomic status (X-axis) and student's test results (Y-axis). The graph shows approximately 500.000 observations. Source: OECD (2004).

The concept negative social heritage was introduced as part of political discussions in social-educational debate by the Swedish researcher Gustav Jonsson in his doctoral thesis from 1967 (Jonsson, 1967). He presented the results from a study of 100 children in a 24-hour care center in Stockholm. He demonstrated that a majority among the placed children came from socially challenged homes and were judged to have inherited the weak aspects, e.g. school attainment tests from their parents and grandparents. Hence, the notification 'negative', since focus was the 'weak' or 'negative' aspects only.

Several Danish researchers within the field of sociology worked in the 1970's with the concept and sharpened the definition in order to make it operational in educational and sociological research, e.g. Erik Jørgen Hansen (2003) and Kuschel & Schulz Jørgensen & (2003). In Hansen (2015) 'unequal chances' is discussed under the caption that socioeconomic background influence's the chance or probability of completing education beyond compulsory schooling: Children from academic background do have a considerable higher probability of obtaining education beyond compulsory schooling compared to children from non-academic homes.

The index ESCS was presented in the PISA 2000 analyses (OECD, 2003; 2004). It was based on five variables: highest occupational status, highest educational level (counted as years of education), a measure of wealth and two measures of cultural possessions and home educational resources. The following PISA studies in 2003 and 2006 restricted the number of variables and made use of only highest parental education, highest parental occupation, and number of home possessions including books in the home (OECD, 2005, 2009). This is, in fact, the variables used earlier, e.g. by Hansen and Kuschel & Schulz Jørgensen, to assess the socioeconomic level.

Both PISA, TIMSS and PIRLS have included an old well-known proxy for ESCS, derived from the three variables: Number of books at home. This variable responded by the student in the questionnaire is recoded into a four-level variable (0-25, 26-100,100-500, more than 500). It has been known for decades to be a proxy for the correlation between students' social heritage and student performance (Engzell, 2019).

The statistical method used for the analyses of the negative social heritage is log linear modelling, more specifically graphical modelling (Edwards, 2000), analyzing higher order contingency tables which are constituted by grouped TIMSS variables from students, parents and teachers as predictors and student TIMSS scores as object variables. The analysis is multidimensional comprising as many variables as possible, and a saturated model for all variables included are formulated with simple main effects and stepwise higher order interaction parameters. The reason for an upper limit of variables to be handled is the fact, that the observations in the cells of the contingency tables become very sparse. A hierarchical step-down procedure is undertaken, trying to successively eliminate (i.e. testing the interaction parameter equal to zero) higher order interaction effects using conditional independence tests (exact chi-square).

The simplest possible example showing the principle with conditional independence graph is the following, where marginal chi-squares reject a hypothesis that being late for Martin (yes/no) is independent of being late by Norman (yes/no). A low significance probability has the consequence that the two events are connected by a line, as illustrated in figure 2:

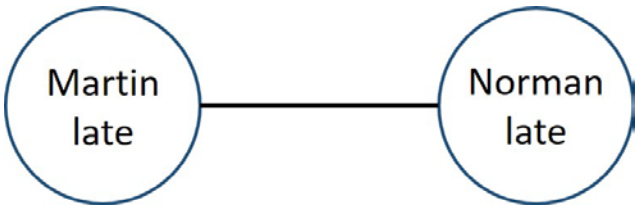

Figure 2: Graphic illustration of conditional dependence between two events.

Now, by including traffic information for Martin and Norman it shows, that by travelling by the same train they are subject to a destiny determined outside their own will. If information as to the regular running of the train due to strike/no strike is included as another variable, the conditional tests for independence runs quite differently, as illustrated in figure 3.

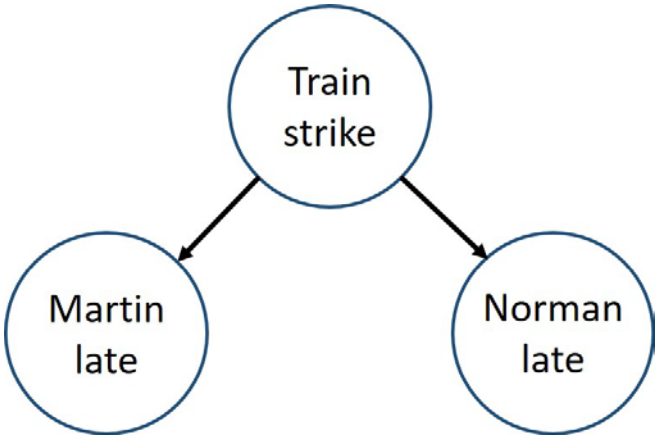

Figure 3: Graphical illustration of independence between two events.

It is now clear, that Martin and Norman being late are events, which are conditional independent – given train strike (yes/no). It is the train strike, that is a confounder behind the simple marginal correlation in the first graph – and the line between Martin late and Norman late is removed (insignificant exact conditional chi-square).
The key variables, which was entered into the separate log linear analyses, are:

Student performance	Mathematics, science
Student background	Gender, language, time of birth, resources
Student theme	Breakfast, absence, bullying, sense of school belonging, like learning mathematics/science, engagement, confidence
School variables	Instructional time, home work in school, school size, urbanization, private/public school, shortage of resource, capacity teaching, discipline, security
Teacher variables	Gender, age, subject specialization, language, resources, safe and orderly, teaching conditions, success, satisfaction, challenges, frequency teaching
Parent variables	Gender, language, resources, assistance and interest in students homework, impression of school, attitude towards mathematics/ science, literacy activity, numeracy activity, literacy tasks, numeracy tasks

Table 1: Overview of key variables used from the Danish part of the TIMSS 2015 study.

Results and discussion

The negative social heritage is strong in Denmark, in fact around 13% estimated from the relation

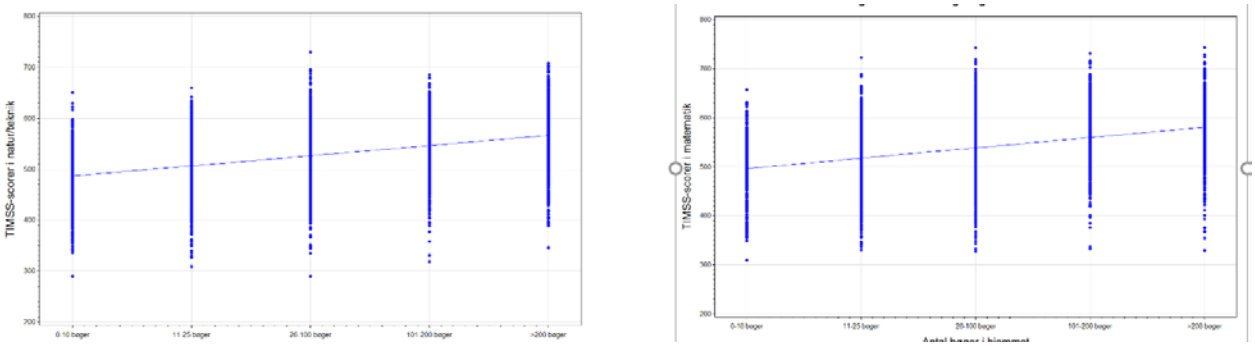

Figure 5: Relation between the ESCS proxy 'Number of books at home' and students score in Science and Mathematics respectively. TIMSS 2015, Danish students.

shown in Figure 4, where X-axes are 'number of books', the proxy for ESCS, and Y-axes are TIMSS 2015 performances in mathematics (left) and science (right) (Allerup et al, 2016). In the log-linear analyses to follow, an extended proxy for ESCS has, however been applied[1]. By this, the analysis utilizes a maximum of information within the TIMSS frame work (see TIMSS 2015, exhibit 4.1 Home Resources for Learning) in order to assess an adequate measure of the student's socioeconomic background. This measure is mapped as 'Resource' in the graphs.

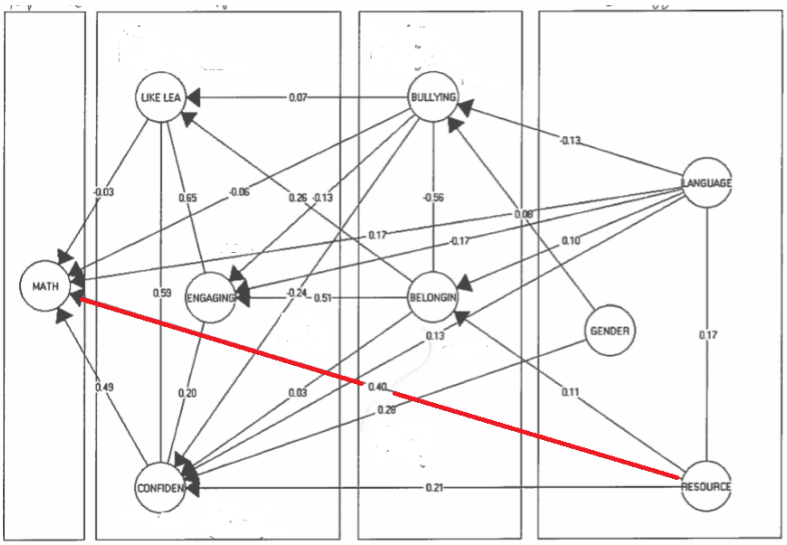

Figure 4: Graphic illustration of contingent dependency between variables correlating with student performance in mathematics. Source: TIMSS 2015, Danish students.

[1] Number of BOOKS, number of study POSSESSIONS, number of children's BOOKS, Highest level of education (parents) EDUCATION, Highest level of occupation (parents) WORK.

Figure 5 includes the variables: student gender, language, resources, bullying, sense of school belonging, like learning mathematics, engagement teaching in mathematics, confidence in mathematics, and, finally, performance in mathematics.

Following the idea behind the graphs for Norman and Martin the aim is to eliminate/replace the red line connecting Resource (lower right corner) and Math (left box) with other arrows pointing to math performance. The strongly significant red relation (correlation= gamma =0.40 from contingency tables) however, and many other significant arrows[2] points to the fact, that this cannot be achieved!

All analyses of this kind result in more or less the same picture. The discouraging fact is that whatever variables are included in the graphs besides the main variables Resource and mathematics – or science – performance – the basic gamma-correlation remains around 0.40, indicating significant connecting between the two variables, e.g. the following supplementary example in Figure 6:

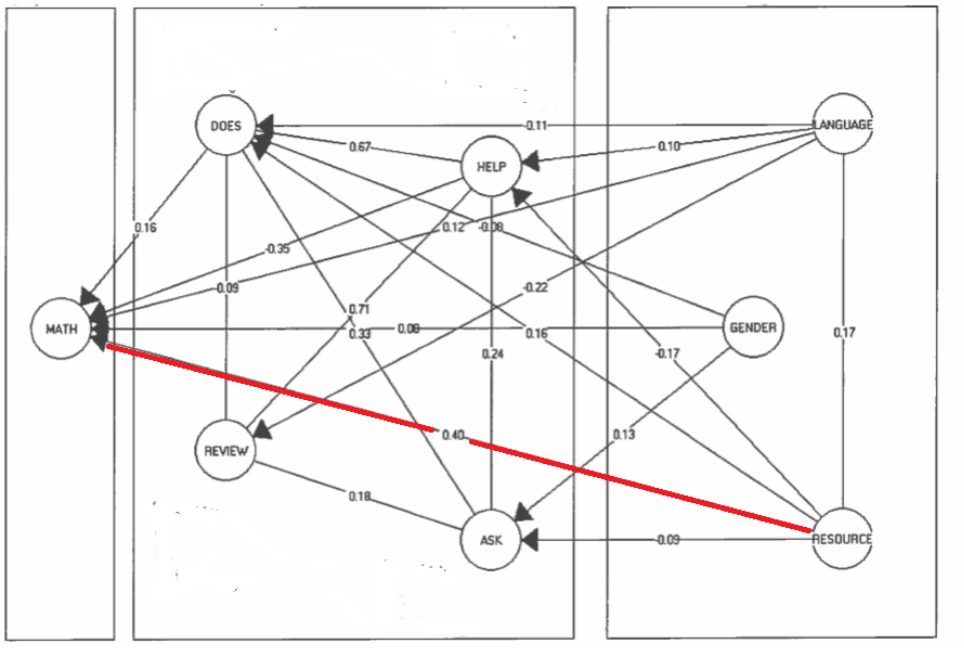

Figure 6: Graphic illustration of contingent dependency between variables correlating with student performance in mathematics. Source: TIMSS 2015, Danish students.

Where the variables are related to parents: Gender, language, student frequency of homework, ask student about completing homework, help student with homework, review students homework.

[2] Only significantly correlating variables are shown by connecting arrows

Still, along other significant connections the main variables are inter connected by a significant gamma correlation =0.40.

Conclusions

The aim of the paper was to use graphical Modelling techniques on higher dimensional contingency tables to analyze the connection between student achievements in mathematics and science and the student's socioeconomic background (ESCS) using Danish students from the TIMSS 2015 data. More specifically the objective was to replace a significant marginal correlation between SA and ESCS with confounders, defined trough conditional independence analysis among the many contextual variables for students, teachers, parents and school principals in the TIMSS data.

The results however showed that the correlation between student achievement and ESCS *persisted, irrespective of whatever variables were introduced as possible confounders behind the correlation.* It is difficult to accept a factual interpretation of the ESCS (or used proxies) as relevant predictor of the student's achievement. However, the results show, that the relation are not confounded by effects on what the student brings into school in terms of e.g. language or time of birth (student background), how the student meets the school in terms of e.g. absenteeism, bullying and self-beliefs (student theme). Selection issues in relation to the school (school variables), or teacher bias due to elements such as qualifications or job satisfaction (teacher variables) are not at play either. Possible relations between students and their parents as for instance parents' interest or support to the child were also not confounding the relationship (parent variables).

The results here are thus in line with other studies (see e.g. Gustafsson, Nielsen & Hansen, 2018; Pokropek, Borgonovi & Jakubowski, 2015) looking at the relation between ESCS and academic outcome for older students.

References

Allerup, P., Belling, M. N., Kirkegaard, S. N., Stafseth, V. T., & Torre, A. (2016). *Danske 4.-klasseelever i TIMSS 2015: en international og national undersøgelse af matematik- og natur/teknologikompetence i 4. klasse* (1. ed.). Fjerritslev: Forlag1.dk.

Edwards, D. (2012). *Introduction to graphical modelling*. New York: Springer Science & Business Media. https://doi.org/10.1007/978-1-4612-0493-0

Engzell, P. (2019). What Do Books in the Home Proxy For? A Cautionary Tale. *Sociological Methods & Research, 0*(0). https://doi.org/10.1177/0049124119826143

Gustafsson, J.-E., Nilsen, T., & Hansen, K. Y. (2018). School characteristics moderating the relation between student socio-economic status and mathematics achievement in grade 8. Evidence from 50 countries in TIMSS 2011. *Studies in Educational Evaluation, 57*, 16-30. https://doi.org/10.1016/j.stueduc.2016.09.004

Hansen, E.J. (2003): *Uddannelsessystemerne i sociologisk perspektiv*. København: Hans Reitzels Forlag.

Hansen, E.J. (2015) *Social Mobilitet*. København: Hans Reitzels Forlag.

Jonsson, G. (1967). Delinquent boys, their parents and grandparents. *Acta psychiatrica Scandinavica, 42*(s195), 1-256. https://www.doi.org/10.1111/j.1600-0447.1967.tb10180.x

Kuschel, R., & Schultz Jørgensen, P. (2003). Socialpsykologi: en introduktion. København: Frydenlund.

OECD. (2003). Programme for International Student Assessment (PISA) (R. Adams & M. Wu Eds.). Paris: OECD. https://doi.org/10.1787/9789264199521-en

OECD. (2004). *Learning for Tomorrow's World*. Paris: OECD Publishing. https://doi.org/10.1787/9789264006416-en

OECD. (2005). *PISA 2003 Technical Report*. Paris: OECD. https://doi.org/10.1787/9789264010543-en

OECD 2009 OECD. (2009). *PISA 2006 Technical Report*. Paris: OECD. https://doi.org/10.1787/9789264048096-en

Pokropek, A., Borgonovi, F., & Jakubowski, M. (2015). Socio-economic disparities in academic achievement: a comparative analysis of mechanisms and pathways. *Learning and Individual Differences, 42*, 10-18. https://doi.org/10.1016/j.lindif.2015.07.011

TIMSS (2015) *TIMSS 2015, exhibit 4.1 Home Resources for Learning*. Boston: TIMSS & PIRLS International Study Center. http://timss2015.org/download-center/